

CLAIMS

What is claimed is:

1. A computerized system for augmenting data from a source database with data from a reference database to generate an augmented database that can be used for predictive modeling, comprising:
 - a) a source database comprising structured data;
 - b) a reference database having reference data;
 - c) a locator component configured to use the structured data to locate reference data in the reference database suitable for association with the source database;
 - d) an analyzer component configured to process the reference data into a set of descriptors and associating the descriptors to the source data to form an augmented database;
 - e) a predictive modeling component configured to classify behavior with the augmented database; and
 - f) a data mining component configured to conduct searches of data in the augmented database.
2. The computerized system of Claim 1, wherein the source database contains financial transaction data.
3. The computerized system of Claim 1, wherein the source database contains telephone call detail records, and wherein the reference database contains business indices and telephone directories augmented by public information on merchants and service providers.
4. The computerized system of Claim 1, wherein the source database contains investment transactions and the reference database contains public information regarding companies, mutual funds and/or other investment interests.

5. The computerized system of Claim 1, wherein the source database contains insurance transactions, and wherein the reference database contains information regarding insurance products, claims and/or insurance evaluations.
6. The computerized system of Claim 1, wherein the source database contains product inventories, and wherein the reference database contains information describing products.
7. The computerized system of Claim 1, wherein the source database contains Internet browser view transactions, and wherein the reference database contains the Internet pages of the browser view transactions.
8. The computerized system of Claim 1, wherein the source database contains retail transactions at an individual product level, and wherein the reference database contains product information from catalogs.
9. The computerized system of Claim 2, wherein the structured data comprises at least a name or identifier corresponding to a merchant, product and/or service.
10. The computerized system of Claim 1, wherein the reference database contains data in an unstructured format.
11. The computerized system of Claim 10, wherein the reference database comprises a public database such as the Internet.
12. The computerized system of Claim 11, wherein the locator component locates electronic pages on the Internet related to merchant, product and/or service identified of the structured data in the source database.
13. The computerized system of Claim 12, wherein the locator component includes a spider module that searches for embedded links, keywords and/or references in the text found at the located electronic pages.

14. The computerized system of Claim 12, wherein the locator component retrieves the natural language text from the located electronic pages.
15. The computerized system of Claim 14, wherein the processing of reference data in the reference database is accomplished by reducing the natural language text to a set of weighted keywords.
16. The computerized system of Claim 12, wherein the locator component validates the located electronic pages using zip code and/or Standard Industry Code (SIC) information stored in the source database.
17. The computerized system of Claim 1, wherein the predictive modeling module uses one or more of the following methodologies: model-based regression, non-parametric regression (*e.g.*, neural networks), Bayesian inference, hidden Markov models, fuzzy logic models, evolutionary models, or decision trees.
18. The computerized system of Claim 1, wherein the source database comprises account based transactional records and the analyzer component aggregates the data from the source database and its associated reference data by reference to an account field.
19. The system as defined in Claim 18, wherein the association of unstructured data from the reference database is delivered through a predictive statistical model built from known historic outcomes associated with records within the source database.

20. A computerized system for augmenting data from a source database with data from a reference database to generate an augmented database that can be used for predictive modeling, comprising:
- a) a source database comprising a plurality of transaction data records with each transaction data record having at least one field identifying a merchant, product and/or service;
 - b) a merchant identifier database comprising a plurality of reference addresses and value description identifiers for merchants, products and/or services;
 - c) a reference database;
 - d) an address locating module configured to search the reference database to locate references for merchants, products and/or services identified in the source database;
 - e) an account description database;
 - f) a transaction augmentation module, configured to attach the value description of a particular merchant, product and/or service to the transaction data records and store the resulting combined record in the account description database; and
 - g) a merchant analysis builder module configured to condense the references provided by the address locating module into a value description and store the value description in the merchant identifier database.
21. The system of Claim 20, further comprising an account descriptor builder module configured to generate descriptive account records from the merchant identifier database and the source database.
22. The system of Claim 20, further comprising a lexicographic database configured to index value description identifiers to keywords.

23. The system of Claim 20, wherein the reference database comprises the Internet.
24. The system of Claim 20, further comprising a predictive modeling module configured to predict future behavior of accounts, merchants, or other entities, using data from the account description database.
25. The system of Claim 20, further comprising a data mining search engine configured to conduct keyword searches of the account description database to identify accounts, merchants, or products.
26. A computerized method of augmenting data from a source database with data from a reference database, the method comprising:
- a) retrieving at least one data record recording an event from the source database;
 - b) identifying a field in the data record that specifies an entity;
 - c) locating reference data from the reference database that describes the entity specified by the entity field;
 - d) processing the reference data to form a set of keyword descriptors describing the entity;
 - e) augmenting the data record with the keyword descriptors to generate an augmented data record describing the entity;
 - f) building an account descriptor database that includes at least one data record that correlates the at least one event with the description of the entity from the augmented data record; and
 - g) searching the account descriptor database for selected data records that meet a desired criteria.
27. The method of Claim 26, wherein the locating reference data includes locating data in an unstructured database.

28. The method of Claim 26, wherein the reference database includes at least a portion of the Internet.
29. The method of Claim 28, wherein the locating reference data includes locating electronic pages using the entity specified in the at least one data record.
30. The method of Claim 29, wherein locating reference data further spidering for additional electronic pages cited within the located electronic pages.
31. The method of Claim 26, wherein locating reference data includes reducing natural language text to keyword descriptors.
32. The method of Claim 26, further comprising validating the located reference data using data from the at least one data record.
33. The method of Claim 26, further comprising storing the augmented data record in a merchant database.
34. A method of augmenting structured data stored in a source database with unstructured data stored in a reference database, comprising:
 - a) reading a data record from the source database;
 - b) searching the reference database for information describing the data record;
 - c) condensing the information describing the data record into at least one keyword description;
 - d) augmenting the data record with the keyword description.
35. The method of Claim 34, wherein the reference database comprises the Internet.

36. The method of Claim 35, wherein the data record contains at least a merchant name or identifier.
37. The method of Claim 36, wherein searching the reference database further includes locating electronic pages related to the merchant identified in the data record.
38. The method of Claim 37, wherein searching the reference database further includes retrieving the natural language text from the located electronic pages.
39. The method of Claim 38, wherein condensing the information comprises reducing the natural language text to at least one weighted keyword.
40. A computerized system associating unstructured data in a reference database with structured data in a source database, comprising:
- a) means for reading a data record from the source database;
 - b) means for searching the reference database for information describing the data record;
 - c) means for condensing the information describing the data record into at least one keyword description; and
 - d) means for augmenting the data record with the keyword description.
41. The computerized system of Claim 40, wherein the reference database comprises the Internet.
42. The computerized system of Claim 41, wherein the data record contains at least a merchant name or identifier.
43. The computerized system of Claim 42, wherein the means for searching the reference database includes means for locating electronic pages related to the merchant identified in the data record.

44. A structured database for use with a database mining search engine conducting searches for behavioral driven characteristics, said structured database containing keyword descriptors describing a merchant obtained by reducing information about the merchant from a reference database comprising unstructured data.
- 5
45. A method of generating a behavior driven targeted marketing list, comprising:
- a) obtaining a plurality of financial transaction records associated with a plurality of individuals and at least one merchant;
 - b) identifying the merchant involved in the financial transactions;
 - c) searching a reference database of unstructured data for information about each of the at least one merchant;
 - d) condensing the information into a list of weighted keywords that describes each of the at least one merchant;
 - e) associating the weighted keywords with the financial transaction records;
 - f) generating a profile of each of the individuals using the weighted keywords describing the at least one merchant, wherein the individuals performed a plurality of financial transactions; and
 - g) searching the individual profiles to identify targeted individuals who exhibit a desired behavioral history.
- 5
- 10
- 15

46. A computerized system for augmenting data from a source database with data from a reference database to generate an augmented database that can be used for predictive modeling, comprising:
- a) a source database comprising structured data;
 - b) a reference database having reference data;
 - c) a locator component configured to use the structured data to locate reference data in the reference database suitable for association with the source database; and
 - d) an analyzer component configured to process the reference data into a set of descriptors and associating the descriptors to the source data to form an augmented database;
47. The computerized system of Claim 46, further including a predictive modeling component configured to classify behavior with the augmented database; and
48. The computerized system of Claim 46, further including a data mining component configured to conduct searches of data in the augmented database.
49. The computerized system of Claim 46, wherein the source database contains financial transaction data.
50. The computerized system of Claim 47, wherein the structured data comprises at least a name or identifier corresponding to a merchant, product and/or service.
51. The computerized system of Claim 46, wherein the reference database contains data in an unstructured format.
52. The computerized system of Claim 51, wherein the reference database comprises a public database such as the Internet.

53. The computerized system of Claim 52, wherein the locator component locates electronic pages on the Internet related to merchant, product and/or service identified of the structured data in the source database.
54. The computerized system of Claim 53, wherein the locator component includes a spider module that searches for embedded links, keywords and/or references in the text found at the located electronic pages.
55. The computerized system of Claim 53, wherein the locator component retrieves the natural language text from the located electronic pages.
56. The computerized system of Claim 55, wherein the processing of reference data in the reference database is accomplished by reducing the natural language text to a set of weighted keywords.
57. The computerized system of Claim 53, wherein the locator component validates the located electronic pages using zip code and/or Standard Industry Code (SIC) information stored in the source database.
58. The computerized system of Claim 46, wherein the predictive modeling module uses one or more of the following methodologies: model-based regression, non-parametric regression (e.g., neural networks), Bayesian inference, hidden Markov models, fuzzy logic models, evolutionary models, or decision trees.
59. The computerized system of Claim 46, wherein the source database comprises account based transactional records and the analyzer component aggregates the data from the source database and its associated reference data by reference to an account field.

60. The system as defined in Claim 59, wherein the association of unstructured data from the reference database is delivered through a predictive statistical model built from known historic outcomes associated with records within the source database.